International Journal of Pattern Recognition and Artificial Intelligence


# Video-Based Classification of Driver Behaviour using a Hierarchal Classification System with Multiple Features

Chao Yan*
*Department of Computer Science and Software Engineering, Xian Jiaotong-Liverpool University, SIP, Suzhou 215123, China*
*choise.yan@163.com*

Frans Coenen
*Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK*
*Coenen@liverpool.ac.uk*

Yong Yue
*Department of Computer Science and Software Engineering, Xian Jiaotong-Liverpool University, SIP, Suzhou 215123, China*
*yong.yue@xjtlu.edu.cn*

Xiaosong Yang
*National Centre for Computer Animation, Bournemouth University, Bournemouth BH12 5BB, UK*
*xyang@bournemouth.ac.uk*

Bailing Zhang
*Department of Computer Science and Software Engineering, Xian Jiaotong-Liverpool University, SIP, Suzhou 215123, China*
*bailing.zhang@xjtlu.edu.cn*

Driver fatigue and inattention have long been recognised as one of the main contributing factors in traffic accidents. Therefore, the development of intelligent driver assistance systems, which provide automatic monitoring of driver's vigilance, is an urgent and challenging task. This paper presents a novel system for video-based driver behaviour recognition. The fundamental idea is to monitor driver's hand movements and to use these as predictors for safe/unsafe driver behaviour. In comparison to previous work, the proposed method utilises hierarchal classification and treats driver behaviour in terms of a spatio-temporal reference framework as opposed to a static image. The Approach was verified using the Southeast University Driving-Posture Dataset, a dataset comprised of video clips covering aspects of driving such as: normal driving, responding to a cell phone call, eating and smoking. After pre-processing for illumination variations and motion sequence segmentation, eight classes of behaviour were identified. The overall prediction accuracy obtained using the proposed approach was 89.62% when using a hierarchical classification approach. The proposed approach was able to clearly identify

*Corresponding author. Tel:+86 512 88161502.

two dangerous driver behaviours, *Responding to a cellphone call* and *Eating*, with an overall recognition rate of 91.87%.

*Keywords*: Driver behaviour recognition; Driving assistance system; Gait energy image; Hierarchal classification.

## 1. Introduction

Unsafe and dangerous driving accounts for the death of more than one million lives and over 50 million serious injuries worldwide each year [37]. The U.S. National Highway Traffic Safety Administration (NHTSA) data indicates that 1.6 million nonfatal injuries, and 40 thousands fatalities, resulted from traffic accidents in 2012, with up to 80% of them due to driver inattention [38]. In Europe, up to 20% of accidents are caused by driver drowsiness. Moreover, it [39] was estimated that the worldwide vehicle population would increase to 1.2 billion in 2014. With the ever-growing traffic density, the number of road accidents is anticipated to further increase. Finding solutions to reduce road accidents and improve traffic safety has become a top-priority for many government agencies and automobile manufactures alike.

Statistics show that one of the leading causes of fatal or injury-causing traffic accidents is the diminishment of the driver's vigilance level. The main contributing factors may either be fatigue or distractions. Scientific research has been conducted to estimate the level of sleep deprivation in relation to traffic accidents[3,25]. The development of Intelligent Driver Assistance Systems (IDAS) ,that continuously monitor, not just the surrounding environment and vehicle state, but also driver behaviours, have attracted increasing worldwide attention[48]. IDAS are seen to be particularly relevant with respect to long-distance drivers as they often drive alone. Usage of IDAS that 'flag' important information outside of a vehicle, such as driving lane indicators and traffic signs, have been shown to increase driver alertness[18,32]. However, automatic detection and warning of driver fatigue and distraction level is considered to be of equal importance with respect to road accidents prevention. Other than for reasons of road safety enhancement, there are also commercial reasons for fitting driver alertness monitoring systems, particularly with respect to truck and bus fleet managers.

Most existing vision-based methods (which will be reviewed in section 2) that detect dangerous driver behaviours including fatigue and visual distraction (e.g., looking away from the roadway), focus on examining facial visual features on the eyes and mouth. Analyzing the state of eyes and mouth can provide observable cues for the detection process, which requires specially designed cameras and the accurately eye localisation algorithm. Meanwhile, other kinds of driver manual distraction behaviour including driver's hands off the wheel, responding to a ringing cell phone,and manually adjusting the radio volume, are difficult to analyse through driver's face character. An alternative way to recognise driver manual distraction behaviour is analysing the driver body posture including the position of arms, hands and feet. However, most previously approaches that analyse the driver body posture regard it as a static image classification problem and therefore classify the posture pattern frame by frame. Methods under such framework are not sufficient to distinguish between classes of behaviour types because of similar postures existing in different driver behaviours. We argue that the driver behaviour is a space-time human activity and should be analysed as time series posture sequence. In this paper, a video camera-based system to monitor driver manual distraction behaviour and distinguish between safe and unsafe driver behaviours, which operates according to the analysis of hand movments and usage, is proposed. This entails a number of challenges namely: (i)motion detection and segmentation, (ii)motion representation, and (iii)the classification of the hand gestures. For this purpose, unsafe hand movements and usage include: smoking, eating, using a cell phone and adjusting the controls of the dashboard while driving. A further challenge is the nature of the required video data pre-processing to compensate for noise and illumination variation.

Specifically, in the proposed video-based driver behaviour recognition system, raw video data was first pre-processed to compensate for illumination changes to improve the performance of motion detection. The pre-processing procedure uses a proposed two stage intensity normalisation technique to minimise the influence from illumination variation. Next, the processed video data was segmented into video clips based on the existence of motion. In this system, then the motion clips were then represented using Gait Energy Image [24] and Pyramid histogram of gradient [5] to reduce data dimension. Finally, a hierarchal classification system is applied to improve the recognition performance. The proposed approach was tested on the Southeast University Driving-Posture Dataset (SEU dataset). It includes activities of normal driving, responding to a cell phone call, eating and smoking.

Given the above, the contributions of the paper are as follows:

(1) A view-based spatio-temporal template approach to represent driving video sequences and that (as will be evidenced later in this paper)archived competitive performance. Contrary to many previously published work, this paper argues that driver behaviour analysis is better treated as a spatio-temporal problem as opposed to a static images analysis problem; as driver behaviour analysis is a space-time human activity. It is argued that usage of static images is not sufficient to distinguish between classes of behaviour types and that this can only be done by considering a sequence of images (video frames).

(2) To minimise the influence from illumination variations, a two stage intensity normalisation preprocessing technique is proposed. The first stage comprises a moving average method that smoothens the intensity variation caused by periodic lighting change. The second stage comprises application of the three frame difference method[19] to detect motion. For the task of motion detection and segmentation in video, it is found that the proposed two-stage pre-processing technique performs well in context of compensating for noise and illumination variation in video data.

(3) A hierarchal classification system for driver behaviour recognition, which considers different sets of features at different levels. Hierarchical classification is specifically intended for data where the features of interest can be arranged in a hierarchical manner. As such it offers advantages in terms of learning and representation in comparison to attempts to use "flat" classification techniques for the purpose of classifying hierarchical data[62]. These efficiency gains are realised because only a subset of the complete set of available features is considered at each node in the hierarchy. Hierarchical classification schemes have been applied in many areas [56,43,35]. However, it should be noted here that, to the best knowledge of authors' knowledge, they have not been applied to driver behaviour recognition.

The rest of the paper is organized as follows. Section 2 presents a review of previous work, while Section 3 gives a brief introduction to the SEU driving dataset followed by an overview of our proposed recognition system in Section 4. Section 5 explains the nature of the required preprocessing of the video data especially in the context of illumination variation. Section 6 introduces the driving motion segmentation algorithm and motion representation by Gait Energy Image (GEI) representation. Section 7 gives details of the hierarchal classification system adopted to predict driver behaviour. Section 8 reports the conducted evaluation and the experiment results obtained, this is followed by conclusions presented in Section 9.

## 2. Previous Work

Previous works on vision-based automatic monitoring of unsafe driver behaviours [17] can be categorized into three main streams of activity: (i) gaze and head poise analysis with which to predict driver behaviour and intention, (ii) extraction of fatigue cues from driver facial images and (iii) characterization (in the context of safe versus unsafe driving behaviour) of driver body posture including the positioning of arms, hands and feet. The proposed system presented in this paper can be said to fall into the third stream of activity.

With respect to the first stream of activity. Wahlstrom et al. [51] proposed a mechanism for locating the eyes and pupils in a facial image using skin colour area and then estimating the gaze direction from the relative positions of the eyes and pupils. Of course this approach will not succeed if the driver's head is turned away from the camera. In order to minimize the influence of various illumination and background interferences, infrared cameras were used in the work presented in [26] to estimate the driver face direction, again based on skin colour area analysis. To improve the performance of head pose estimation, in the presence of dramatic changes in illumination, the use of isophote features was introduced in [59]. In [52], video frames were represented using the Fisher face approach and then classified using the nearest neighbor and neural network models. However, the system is driver dependent, which makes it unrealistic in many situations. An integrated system for monitoring driver awareness, based on head pose estimation, was presented in

[34], which include head detection and tracking. A comparative study of the influence that eye gaze and head movement dynamics have on (i)driver behaviour and (ii)intent prediction with respect of lane change manouvers was presented in [18].

The second main stream of research,as noted above, focuses on the extraction or recognition of fatigue cues the driver faces (for example yawning). A method was proposed in [20] to locate and track driver mouth movements with the aid of template matching for face localization and simple image processing for mouth corner detection. In [2], Gabor filtering and Local Binary Pattern (LBP) description were jointly applied to characterize driver yawning. However, experiments were only conducted using a small number of frontal face images. To better describe and classify driver fatigue expression, feature fusion was considered in [60] coupled with the use of a classifier ensemble. In addition to facial fatigue expression, eye blink pattern is another important sign indicative of fatigue (or lack of). There is much reported works along this line. For example, a fuzzy classification system was proposed in [3] to infer the driver's vigilance level by estimating some parameters which characterize eye closure and blink frequency. A probabilistic model was proposed in [25] ,to predict fatigue, based on different visual cues which included eyelid movement.

The third main stream of research, directed at vision-based automatic driver behaviour prediction, centers on the characterization of driver body posture, including arms, hands and feet. For example, a variant of the Iterative Closest Point (ICP) registration algorithm was proposed in [16] to estimate the location and orientation of a driver's limbs, with visual information provided by an infrared Time-of-Flight camera. Driver posture dynamics in 3D was investigated in [47] using a vision-based system. In [11] a camera array system was proposed to track important driver body parts and to analyze driver activities such as steering movements. In [49] an agglomerative clustering and Bayesian eigen-image approach were applied to represent and recognize predefined safe/unsafe driving activities, such as talking on a cellular phone and eating. A modified Histogram of Oriented Gradients (HOG) feature description mechanism coupled with a support vector machine classifier was applied in [12] to discriminate which of the front-row seat occupants was accessing "infotainment" controls. To investigate "pedal error phenomenon" Tran et al. [46] developed a vision based system for driver foot behaviour analysis which featured an optical flow based foot tracking and a Hidden Markov Model (HMM) based approach to characterize temporal foot behaviour.

## 3. The SEU Driving Dataset

To test the proposed driver behaviour recognition approach, the Southeast University Driving-Posture Dataset (SEU dataset) was used. This data was first created by Zhao [60]. Some selected frames from this dataset are shown in Fig.1. Each video included in the dataset was obtained using a side-mounted Logitech C905 CCD camera under day lighting conditions with a resolution of 640x480. Ten male drivers and ten female drivers participated in the creation of the dataset. Each video was

Fig. 1. SEU driving dataset

recorded under normal day light conditions, poor illuminated night time conditions were not considered.

## 4. System Overview

A schematic illustrating the operation of the proposed driver behaviour recognition system is shown in Fig.2. In the figure the directed arcs indicate the next step followed by previous one. The proposed system comprises following five steps:

Step 1 Motion Detection. Contrary to many previously published works, our design treats driver behaviour analysis as time-series motion classification, as opposed to a static images classification problem. We derive feature representation from motion object silhouettes [55,22], which however requires effective motion detection and segmentation if illumination variation exists. In the first step, we pre-process the input video to compensate for noise and illumination variation, using a proposed two stage intensity normalisation preprocessing technique.

Step 2 Motion Segmentation. In this step, the input video stream is temporally segmented into fragments or clips [53], each of which is a motion clip (image sequence) and contains continuous driver movement without pause.

Step 3 Motion Representation. Given an input motion clip, it is represented into four different gray level images using four methods. Each of the extracted gray level images somehow represent the driver motion in clip as the feature. The pyramid histogram of oriented gradients (PHOG) method [5] is applied

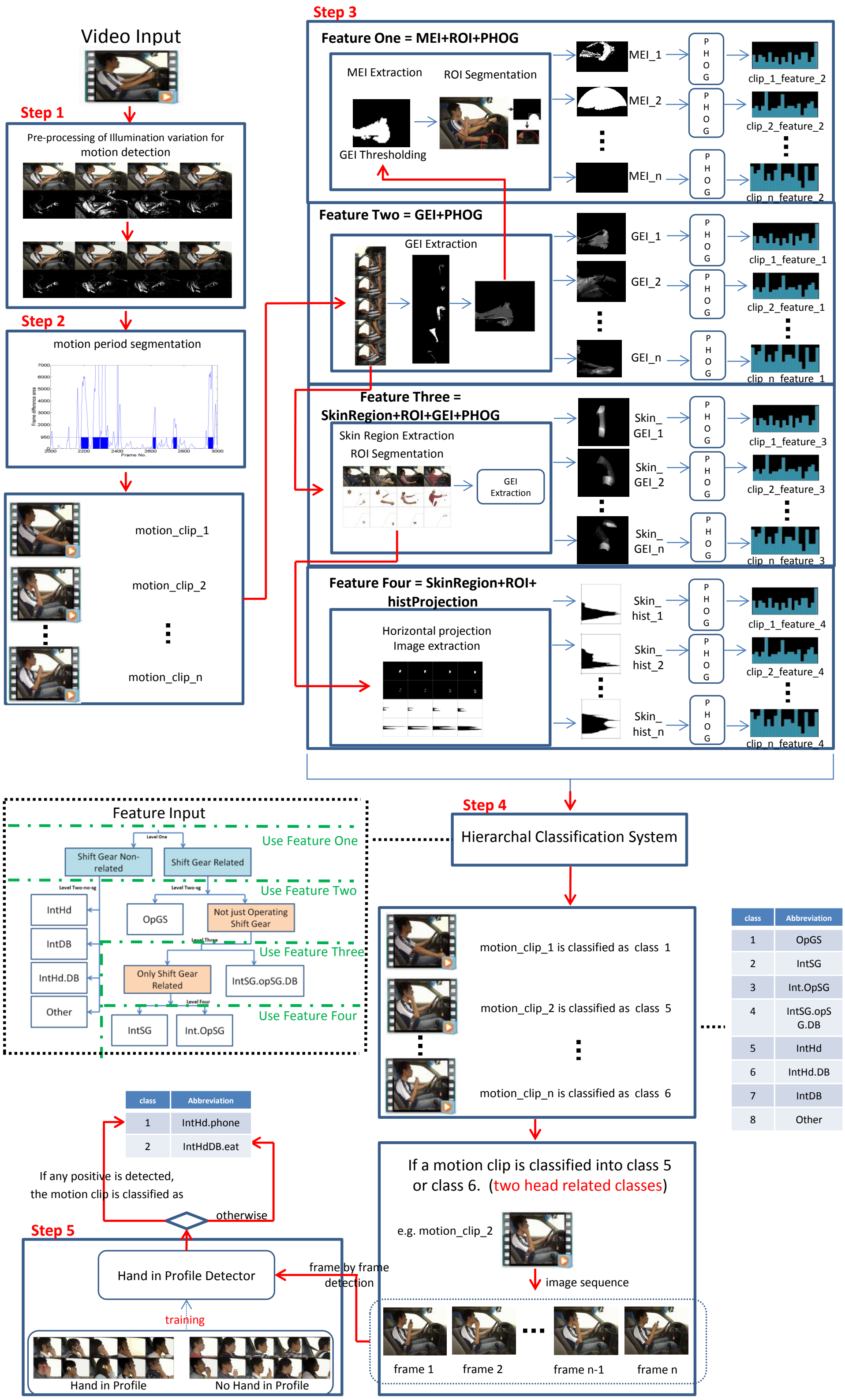



Fig. 2. System overview.

Table 1. Driver's hand movement class definition

| Class | Abbreviation | Description |
|---|---|---|
| 1 | OpGS | The normal operation of the gear shift. |
| 2 | IntSG | Interaction with the gear shift. Thus the movement of the right hand from the steering wheel to the gear shift, or the reverse procedure. |
| 3 | Int.OpSG | Interaction with the gear shift and then operation of the gear shift. It represents compositional behaviour comprising IntSG and OpSG |
| 4 | IntSG.opSG.DB | Interaction with and operation of the gear shift, followed by movement to Dashboard. The class describes the situation where right hand is first used to operate the gear shift, then moves back to the steering wheel and then reaches towards the dashboard. |
| 5 | IntHd | Describes situation where the driver moves his right hand towards or away from his/her head. For example moving food towards the mouth or taking a call by moving a cell phone towards the ear (we call this "head interaction) |
| 6 | IntHd.DB | Interaction between head and dashboard, encompasses IntHd.DB and IntDB |
| 7 | IntDB | Describes situation where the driver moves his right to place something on the dashboard or take something away from the dash board. For example, taking a cigarette from a packet or replacing a cigarette lighter. |
| 8 | Other | Behaviour undefined in the previous seven classes, such as turning of the steering wheel. |

Table 2. Dangerous driver behaviour class definition

| Class | Abbreviation | Description |
|---|---|---|
| 1 | IntHd.phone | Driver takes a cellphone from somewhere, such as dashboard, and place it on the profile of head |
| 2 | IntHdDB.eat | Either eating or smoking a cigarette. |

on the gray level image to further reduce the feature dimension.

Step 4 Hierarchical Classification of Driver Behaviour. In this step, a specially designed hierarchal classification system is used to classify the input motion clip. Different features and classifiers are used in different levels. A Given input motion clip is classified as one of eight kinds of driver's hand movement class, each of which is defined according to the driver's hand movement. (as in the Table 1). From the table it can been seen that the identified eight driver's hand movement classes are defined in terms of the physical position and/or movement of a driver's hand.

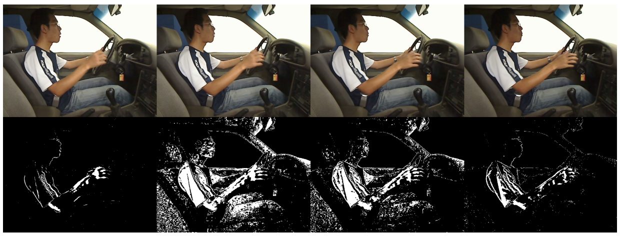
Step 5 Dangerous Driver Behaviour Classification. In the Table 1, IntHd (class 5) and IntHd.DB (class 6) are two head related behaviours. If a motion clip is classified into class 5 or class 6 in previous step, the motion clip is able to indicate that the driver is responding a cellphone, eating or smoking. Therefore, a "Hand in Profile" detector is trained to examine each frame in a class 5 or class 6 motion clip. If "Hand in Profile" is detected in one or more frames in a motion clip, it is classified as responding a cell phone, otherwise, it is eating or smoking. (as in the Table 2)

## 5. Motion Detection

The task of driver behaviour monitoring can be generally studied within the human action recognition framework [53], that is action detection, action segmentation, action representation and action classification. The emphasis of the framework is often on finding good feature representations tolerant of variations in viewpoint, human subject, background, illumination, and so on. One of the common strategies of representing human motion is global description, which regards the visual observation as a whole. Global representation can be derived from motion object silhouettes [55,22] based on effective motion detection and segmentation.

There are three commonly used approaches to detect motion or moving objects, including (i) temporal differencing, (ii) background subtraction, and (iii) optical flow. In the temporal differencing method, the motion is defined as the difference between two consecutive frames. Specifically, a similarity threshold is applied on the subtraction of two consecutive frames to determine whether the frames are different or not [1]. In the background subtraction method, a background image is modeled first as the benchmark image. The motion is identified by calculating the difference between a current frame and the background image [41]. A similarity threshold is applied once again. Both these two methods are able to work well if an appropriate threshold value is applied. However, this is difficult in practice. In addition, the temporal difference approach (and its variants) has the disadvantage of not being able to extract the complete contours of moving objects. In the case of the background subtraction approach a further disadvantage is that it critically relies on precise background modeling, which in turn has a series of open problems. The optical flow method aims to estimate the motion field and merge the motion vectors with similarities. It has been found to work well in the presence of camera motion [44], but requires higher computing capability and is sensitive to noise.

From the above, in action recognition research, temporal difference is often preferred due to its computational efficiency and its consequent potential for usage in real-time applications. However, as noted above choosing a threshold value is a challenging problem. One widely used solution is Otsu's method [1] for selecting a threshold. Otsu's method minimises the intraclass variance of the black and white



(a) Before compensation

(b) After compensation

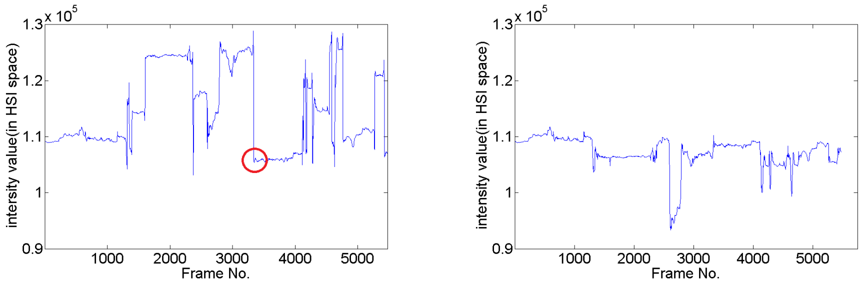
Fig. 3. An example of Negative influence caused using periodic illumination variation and its compensation result

pixels while at the same time being tolerant to slight and slow variation of illumination. The temporal difference motion detection approach, coupled with Otsus threshold selection technique, was thus adopted with respect to the work presented in this paper. However, prior to its application, two kinds of illumination variation found in the SEU dataset had to be taken be addressed, namely: (i)periodic variation, and ii)sudden change. The proposed mechanism for addressing these illumination variation issues are presented in Sub-sections 5.1 and 5.2.

### 5.1. *Periodic Variation*

Periodic illumination variation occurs when a vehicle is passing a sequence of road side objects (such as lamp posts) where by the vehicle under illumination changes in a regular pattern. This type illumination variations thus quasi-periodic and as such is a negative influence on motion detection. This is particularly the case with respect to the temporal differencing approach used with respect to the work presented in this paper because false foreground appears if illumination varies quasi-periodically.

Fig. 3(a) further explains the quasi-periodic illumination variations which arise from the simulated SEU driving dataset. In the figure, the first row comprises an image sequence representing a movement of the right hand reaching towards the gear shift. The second row is the corresponding sequence of frame differences generated by applying temporal difference motion detection (coupled with Otsu's threshold method). The white pixels indicate differences with respected to the previous and consequently are indicative of motion. Obviously, the direct frame differencing re-




(a) Before processing

(b) After processing

Fig. 4. Intensity plot of video 25

sults are too noisy to be proceeded for moving object detection. Such a detriment is caused by the quasi-periodic lighting change, as demonstrated by Fig.4(a), which shows the change of intensity value with time for a specific video sequence. From the figure peaks and troughs in intensity value can be observed. As the video is recorded 30fps, the intensity value jumps roughly about every half a minute.

In order to reduce the influence from the above quasi-periodic lighting change, we proposed an intensity compensation method by smoothing the sharp peaks and valleys. For each frame in a given sequence, we first calculate the difference between the intensity values and the moving average intensity values with respect to a no-motion area. Then we compensate each frame by adding the intensity difference to each pixel in the frame. The process is as follows:

Step 1 For a given video sequence, we calculate the frame difference for each pair of consecutive frames and add these frame differences together. The final aggregated frame difference is thresholded by Otsu's method [1], resulting in a mask for the static pixels. A set of 16 example masks are shown in Fig. 5, with black and white pixels representing motion and no-motion, respectively.

Step 2 The mask from above step 1 is multiplied to its corresponding video frames $I_n$, with $n$ for frame index, to yield the intensity sequences of no-motion area, denoted as $\bar{I}_n$.

Step 3 The moving average of $\bar{I}_n$ is defined as

$$BPI_n = \begin{cases} \bar{I}_n & \text{if } n = 1 \\ (1-a) \times BPI_{n-1} + a \times \bar{I}_n & \text{if } n > 1 \end{cases} \tag{1}$$

where $a$ is a coefficient representing the degree of weighting decrease.

Step 4 The difference $diff_n$ between the $BPI_n$ and $\bar{I}_n$ is is calculated by

$$diff_n = BPI_n - \bar{I}_n \tag{2}$$

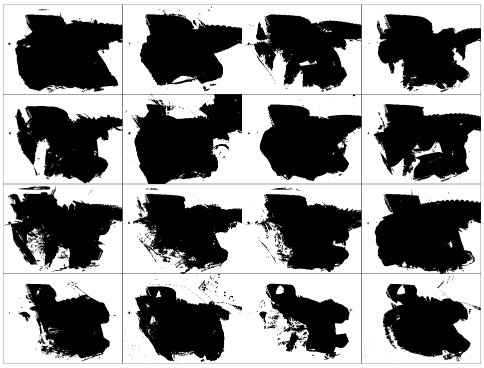


Fig. 5. 16 examples of motion mask, with black pixels representing motion and white pixels representing no-motion

Step 5 Finally, for $n$-th frame $\mathrm{Im}_n$ in the original sequence, the intensity compensated result $\mathrm{Im}'_n$ is given by $\mathrm{Im}_n + \mathrm{diff}_n$

It should be noted that the compensation algorithm is directed specifically at the quasi-periodic illumination variation phenomena. The effect of the above compensation algorithm can be seen by comparing Fig. 4(b) with Fig.4(a). Both figures feature the same video sequence, the first without compensation, and the second with compensation. Noise reduction can clearly be observed from Fig.3(b).

### 5.2. *Sudden Change Variation*

While the influence from quasi-periodic illumination change can be compensated to a large extent by the proposed intensity compensation method, sudden light change remains a problem, which may bring false motion area when the simple temporal difference is applied. In recent years, there have been some exploratory works on the robust moving object detection against fast illumination changes [30,10,13], some of which are extended from temporal difference. For example, a three-frame difference method was proposed in [19], aiming to solve occluded objects detection while alleviating the negative effect from sudden illumination changes. A recent approach [23] uses several temporal reference images to detect moving objects and adapt to sudden illumination change, holes are reduced inside the foreground. However, the detected objects may drag ghost artifacts due to the use of several consecutive frames possibly involving moving objects.

In our works, the three frame difference approach [19] was applied to the intensity compensated sequence to robustly detect moving objects. The approach first applies frame difference to three consecutive frames, and then make an AND operations to the results. Specifically, denote three consecutive frames $f_{k-1}$, $f_k$ and $f_{k+1}$, then two binary images $D_1$ and $D_2$ can be obtained:

$$D_1(x,y) = \begin{cases} 1, & |f_k(x,y) - f_{k-1}(x,y)| \geq T \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$D_2(x,y) = \begin{cases} 1, & |f_{k+1}(x,y) - f_k(x,y)| \geq T \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Then the three difference image is given by as follows:

$$D(x,y) = \begin{cases} 1, & D_1(i,j) \cap D_2(i,j) = 1 \\ 0, & D_1(i,j) \cap D_2(i,j) = 0 \end{cases} \tag{5}$$

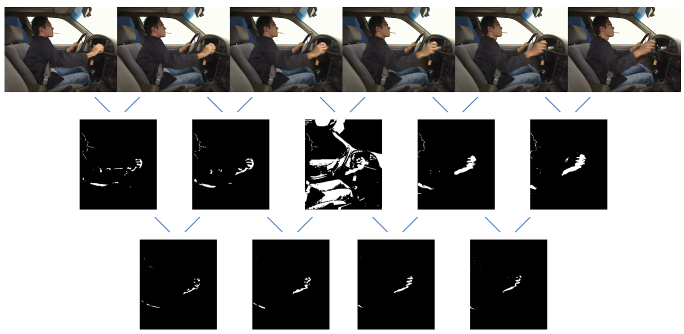


Fig. 6. The first row is the original image sequence after intensity compensation. The second row is the corresponding two consecutive frame differencing image threshold by Otsu's method. The third row is the three frame differencing image corresponded to the second row

.

The performance is shown in Fig. 6, the first row is an original image sequence representing the driver's hand moving back from the dashboard after intensity compensation. There exists an illumination sudden change between the third and forth frame of the first row. The second row is the corresponding two consecutive frame differencing image threshold by Otsus method. The intensity sudden change caused false foreground in the third frame of the second row. By applying three difference method, the three frame differencing image was shown in the third row which proves that the false foreground was reduced.

## 6. Driving Motion Segmentation and Representation

There has been a large body of work that addresses the topic of automatic human action recognition, which focus on the video analysis based on durations and changes of spatial features over time, for example, flow-based iterations [36], motion history image [4], and local interesting points [29]. An implicit assumption on these features, namely, the availability of consecutive frames on a small group of predetermined pixels from which the features are calculated, cannot be made in practice. It remains a challenge to find a generic vocabulary of parts of actions, and the corresponding methods for breaking video streams into the corresponding segments.

Currently, there exists several different kind of methods to temporally segment video streams into fragments or clips [53], including boundary detection [50,42], sliding windows [21,27]and grammar concatenation [7,40]. Among the methods proposed, the boundary detection is relative easy and efficient for the driver behaviour video analysis. Specifically in our approach, motion clips are segmented if there exists a continuity of at least 15 frames with which motion area is over 950 pixels. The two values, i.e., 15 frames and 950 pixels, are from empirical analysis of the SEU datasets. This can be further explained by Fig. 7, which plots the detected motion area in pixels over the frames for the video No.25 of the SEU dataset, showing that six motion clips can be segmented between frames 2000 to 3000. With the simple boundary detection method for video segmentation, 527 motion clips are obtained from 20 raw videos sequence.

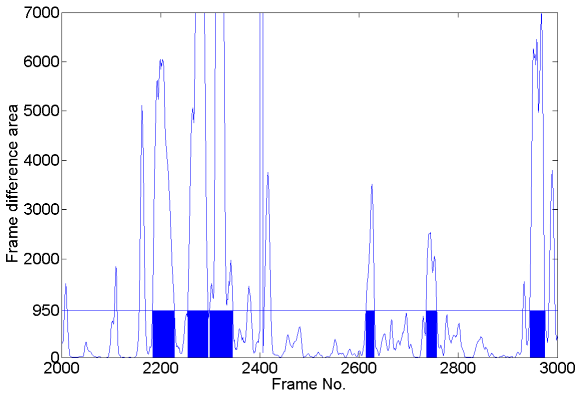



Fig. 7. Motion period segmentation

.

Motion clips segmented from the original video is a sequence of high-dimensional images, which cannot be directly applied for classification. In our earlier work [57], we created a illumination stable driving dataset and manually segmented only four pre-defined motion clips from the video, that is interaction with shift lever, operating the shift lever, interaction with head and interaction with dashboard. The satisfactory performance in experiment has demonstrated the effectiveness of representing motion clips with motion history image (MHI) [4] and pyramid histogram of oriented gradients (PHOG) [5]. Motion history image (MHI) is a view-based temporal approach, which is simple yet robust in the representation of movements and is widely employed in action recognition, motion analysis, and other related applications [6,33,58]. The essence of MHI is to describe motion in the image sequence by representing a pixel intensity as a function of the recency of motion in a sequence, where brighter values correspond to more recent motion. Inspired by MHI, a special motion feature expression approach, termed Gait Energy Image (GEI), was proposed for individual gait recognition [24] and later applied in repetitive human activity classification [63] due to a number of attractive attributes. Recently, some extensions or variants of GEI have been proposed [31,14].

GEI is a simple yet competitive appearance based method that exploits average (i.e., energy) cues as motion features of the whole sequence. With period of gait or other action estimated, GEI can be used to represent the motion with both spatial and temporal information included, and their robustness to specific noises have been proved [54]. GEI is defined as follows:

$$GEI(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \tag{6}$$

where $B_t(x,y)$ is the binary silhouette images at time $t$ in a sequence, $N$ is the number of frames, $t$ is the frame number in the sequence, and x and y are values in the 2D image coordinate.


Fig. 8. Example procedure in extracting gait energy image.

An example procedure of extracting GEI form driver behaviour is illustrated in Fig.8. The first row in left part of the Fig. 8 is an original sequence while the second row in left part of the Fig. 8 is the corresponded silhouette sequence generated from original sequence by the approach described in pre-processing section. The right part of the Fig. 8 is the GEI by averaging the silhouette sequence. From the example gait energy image, it is obvious that higher intensity pixels indicate static areas, while lower intensity pixels highlight dynamic portions of the performed actions.

## 7. Hierarchal Classification of the Driver Behaviour

To alleviate the problems from applying flat classification on overlapping classes, which is obvious for some subclasses defined in Section 4, a commonly applied methodology of hierarchal classification is adopted [56,43,35].

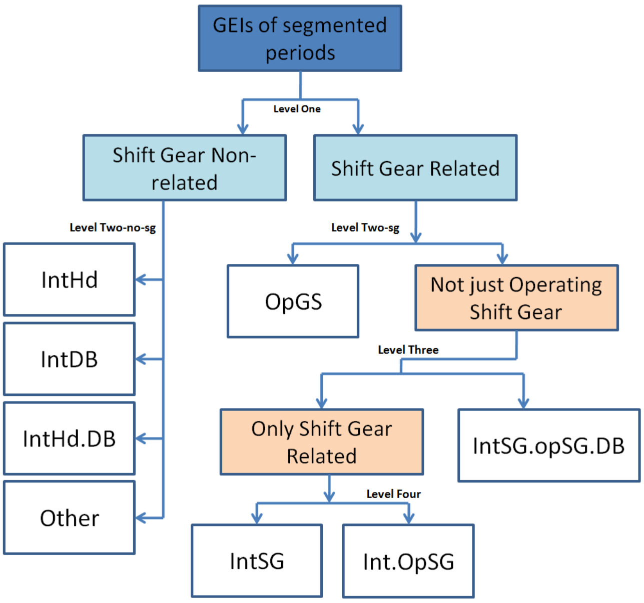



Fig. 9. Hierarchal classification system

.

With the explanatory aid of Fig. 9, a segmented video clip is first classified into *shift gear related* and *shift gear not-related* classes, each of which will be further classified in the next level of the hierarchy. Different regions of interest (ROI) and features can then be exploited for the different subclasses.

### 7.1. *Level One Classification*

We applied SVM classification [28,15] for the first level classes to make a distinction between the *shift gear related* and *shift gear not-related* behaviours. When a driver conducts behaviours including *OpGS* or *IntSG*, the hand will appear in the right bottom in the viewing filed, as indicated by the red circle in Fig. 10. The shift gear related area can then be represented by the motion energy images (MEI) for the two classes, as illustrated by Fig. 11.

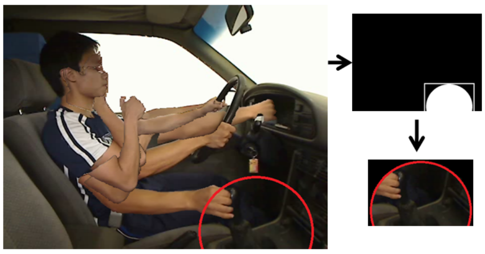


Fig. 10. ROI based on skin region time lapse image

.

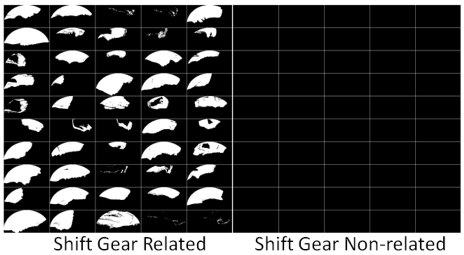



Fig. 11. Two classes in level one of the hierarchal classification system

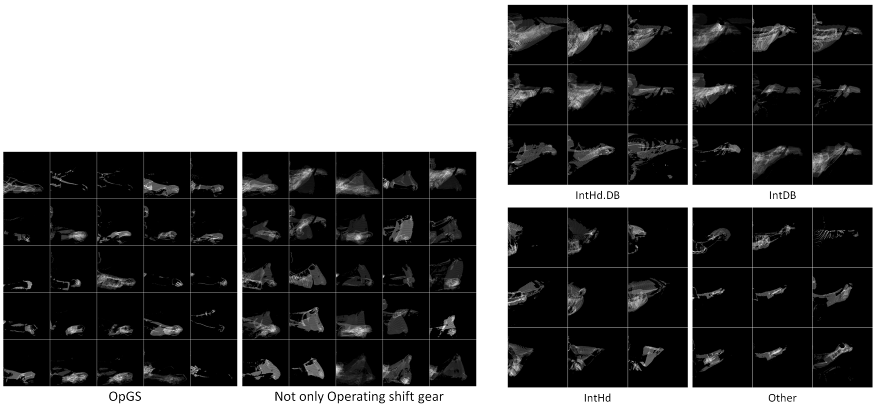
.

### 7.2. *Level Two Classification*

There are two branches in the 2nd level of class hierarchy. The first branch (abbreviated as level two-sg in the figure) categorizes two situations, namely, *OpGS* and *not only operating shift gear*. A random forest classifier [8] is trained to classify the two groups of pattern as shown in Fig. 12(a). The second branch (abbreviated as level two-no-sg in the figure) covers the following four cases: *IntHd*, *IntHd.DB*, *IntHd.DB*, and *Other*, as shown in Fig. 12(b). Similar to the previous discussion, random forest classifier is trained to classify the four groups of GEI.

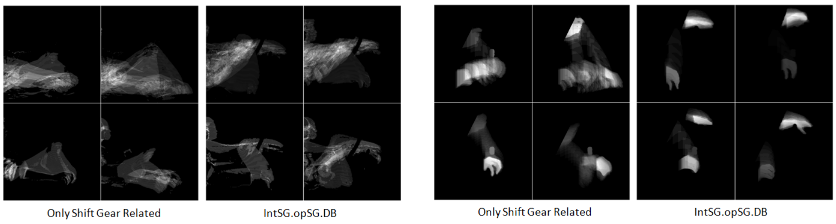
### 7.3. *Level Three Classification*

In the third level of classification hierarchy, two subclasses of the *not only operating shift gear* class are defined, that is *Only shift Gear Realted* and *IntSG.opSG.DB*, as shown in Fig. 13(a). There exists much overlapping if it is represented in the GEI feature space, which makes classification difficult. As the two behaviours are performed by the right hand with motions mainly consisting of moving among shift gear and steering wheel and dashboard, the trajectories of the right hand are easier to distinguish. One possible approach to locate the right hand is by skin-region analysis in a well-defined region of interest (ROI). Specifically, we further extract the right hand skin-region in a ROI for each image of the action sequence, and




(a) Two classes of shift gear related (b) Four classes of no shift gear related

Fig. 12. GEI patterns in level two




(a) Original GEI patterns (b) GEI patterns of right hand in ROI

Fig. 13. GEI patterns in level three

combine them to form a right hand skin-region GEI. There exist many methods for skin region segmentation, for example, difference color space thresholding [9], Gaussian and mixture of Gaussian distributions thresholding method [45]. In this experiment, we simply segment the region of skin based on the following decision rules for the pixel value in YCbCr color space:

$$\begin{cases} 80 \leq Cb \leq 120 \\ 140 \leq Cr \leq 170 \end{cases} \tag{7}$$

Fig. 14 demonstrates the above procedure of locating the right hand skin region in ROI. The first row is four selected frames from the original sequence. The second row is the skin region after applying the above rule corresponding to the first row. As the two classes of behaviours are related to the shift gear region and the dashboard region, the region of interest (ROI) is located at a right trapezoid region of the lower right corner of the frame, which covers the shift gear region and the dashboard region. We only estimate the right hand region in ROI. The third row

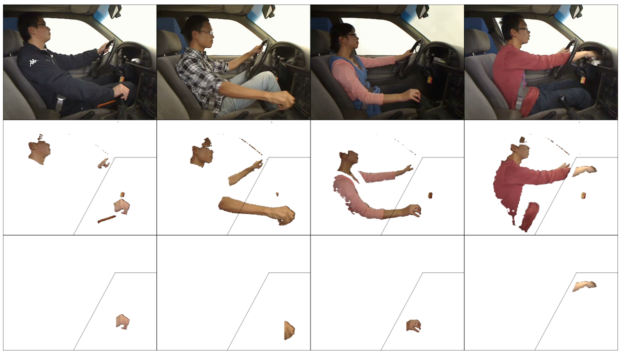


Fig. 14. Locating the right hand skin region in ROI

.

shows the hand region in ROI after connected component analysis and further analysis of the hand area. After locating the right hand skin region in ROI for each frame in the sequence, the right hand region sequence is combined to form another group of GEI, as shown in Fig. 13(b), which is much easier to classify compared to the pattern in Fig. 13(a).

### 7.4. *Level Four Classification*

In the forth level of classification, the class of *only shift gear related* from level three can be further divided into two subclasses, namely *IntSG* and *Int.OpSG*, respectively. However, neither original GEI nor right hand skin region-GEI feature could give a satisfactory separation between these two subclasses. To solve the problem, we propose to exploit features that are more discriminative for hand motions. More specifically, if we summate the vertical projection values on a frame differencing image sequence , a behaviour containing *OpGS* will cause more movement around shift gear which makes larger projection value on the period of vertical axis corresponded to the shift gear area.

Therefore, we calculate the skin region frame differencing sequence and to summate the vertical projection to form a cumulative vertical projection histogram for classification. The detailed steps are as follows:

Step 1 For a given GEI belonging to the class of *only shift gear related*, find its corresponding original frame sequence.

Step 2 Transform the original sequence into a binary image sequence based on hand skin region segmentation proposed in previous subsection.

Step 3 Calculate the frame differencing image sequence from the binary image sequence.

Step 4 For each frame in the sequence, project its binary frame differencing image onto the vertical-axis and get the projection vector.

Step 5 Summate the projection vectors corresponded to each frame to form a vertical projection histogram.

Step 6 Use the histogram to represent a sequence after size normalisation.

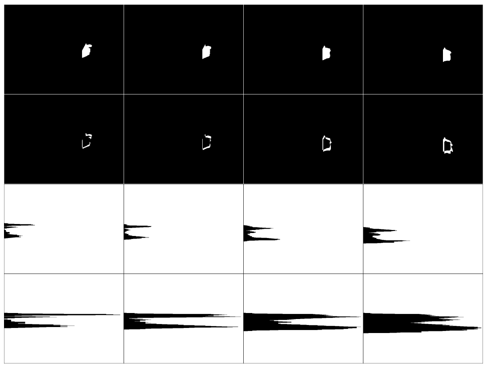


Fig. 15. Right hand skin sequence of video 7 (frame 645–648)and their corresponding horizontal projection image

.

Fig. 15 shows the procedure to generate a horizontal projection histogram. The first row is four consecutive binary frames after right hand skin region segmentation in video 7. The second row corresponds to frame differencing image sequence. The third row shows the horizontal projection histogram corresponding to the frame differencing image in the second row. The forth row is the cumulative horizontal projection histogram. The image of histogram in fourth row and fourth column of Fig. 15 is an example of a cumulative horizontal projection histogram which can be used to represent the motion among the four frames. However, the size of the histogram could be different, we normalise all the histogram to a fixed size.

Fig. 16 shows the normalised horizontal projection histogram of two classes. The sharp peak on the lower side of the histogram of *Int.OpSG* class represents operating the shift gear in the steering room which is the most distinguishing feature by this method.

### 7.5. *Additional Stage Classification on dangerous behaviour*

The segmented driving motion clips are classified into eight classes based on their contents in the previous four level hierarchal classifications. Dangerous driver behaviours, including eating, smoking and responding to a cell phone call, can all be described as the relative motion with reference to the driver's head. Therefore, we perform an additional stage of classification. Specifically, each frame in motion clips from the spatial oriented classes of *IntHd* and *IntHd.DB* will be re-examined and further reclassified into two human perception oriented classes, that is *IntHd.phone*

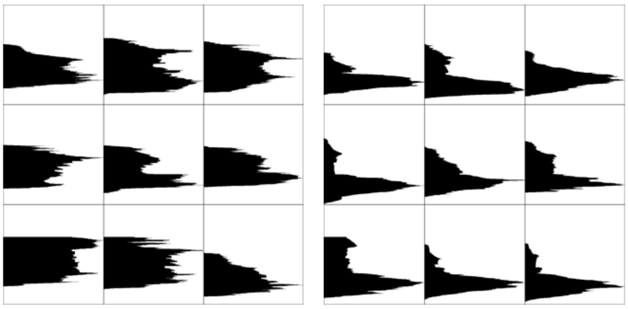



Fig. 16. Normalised horizontal projection histogram of the two classes in level four

.

and *IntHdDB.eat*. In this additional stage, all the frames belonging to classes of *IntHd* and *IntHd.DB* will be further classified into another two classes as shown in Fig. 17.


Fig. 17. Selected frames from the two classes: *no hand in profile* and *hand in profile*

.

The first two rows belong to the class of *no hand in profile* while the bottom two rows belong to the class of *hand in profile*. The PHOG feature is extracted from every frame in every sequence in the *IntHd* class and *IntHd.DB* class. The PHOG feature is used to train and test a k-nearest neighbor (KNN) classifier with good performance. If any frame from the two classes of *IntHd* and *IntHd.DB* is labeled to be hand in profile, the behaviour sequence contains that frame is *IntDd.phone*, otherwise it is *IntHdDB.eat*.

## 8. Experiment

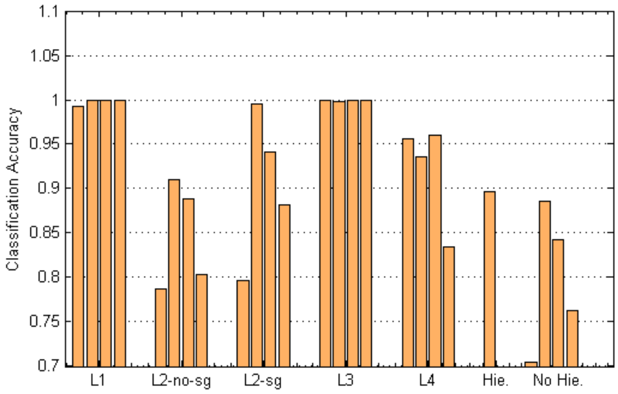
Experiments are carried out to verify the effectiveness of the proposed algorithm on the SEU driving database. This database consists of 20 sequences from 20 drivers conducting eight driver behaviours which have been introduced in section 4. The experiment was conducted on a Dell M6700 workstation with CPU i7 3740QM 2.7GHZ and the proposed algorithm are programmed using MATLAB. In the experiment, 20 videos from the original SEU dataset are first pre-processed to reduce the influence of illumination variation. After that, 527 motion clips are segmented from the original video by the algorithm discussed in section IV. Then eight different classes of motion clips are sent to the hierarchal classification system for training and classification. In order to evaluate the significance of hierarchal system, we also sent the data to a traditional non-hierarchal one-versus-eight classifier for comparison. Finally, we conduct an experiment on additional stage classification for exploring dangerous driver behaviour, one behaviour is *IntHdDB.eat*, the other is *IntDd.phone*. Meanwhile, in each level of the hierarchal system, the non-hierarchal system and the additional stage classification, we compare the classification performance by four commonly used classifiers, that is k-nearest neighour classifier (KNN), random forest classifier (RF), support vector machine classifier (SVM) and multi-layer perceptron classifier (MLP).

### 8.1. *hierarchal and non-hierarchal classification performance*

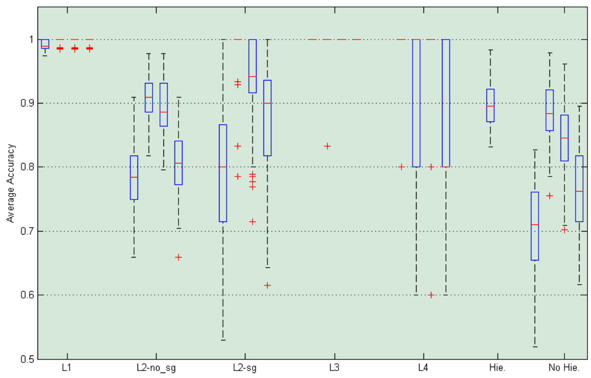
We chose a standard experimental procedure called the holdout approach to verify the driver behaviour recognition system. In the holdout experiment, 10% of the 20 videos, that is 2 videos, are randomly selected as the testing dataset, while the remaining 18 videos are used as the training dataset. The bar plot and box plot of average accuracy results from 100 runs are shown in Fig. 18(a) and Fig. 18(b), respectively.

The ticks in the vertical axis represent level one classification (abbreviated as L1), level two no-shift gear related classification (abbreviated as L2-no-sg), level two shift gear related classification (abbreviated as L2-sg), level three classification (abbreviated as L3), level four classification (abbreviated as L4), hierarchal classification (abbreviated as Hie.), and non-hierarchal classification (abbreviated as No Hie.), respectively. Each tick except Hie. corresponds to one of the four classifier performances(that is, KNN, RF, SVM and MLP, respectively. Table 3 is the numerical results of the bar plot in Fig. 18(a). Based on the performance shown in Table 3, we chose RF in the previous two levels and SVM in last two levels to form the hierarchal classification system, and the final classification accuracy is 89.62%. It has a 1.05% improvement compared to the non-hierarchal classification result of 88.57% which only applies GEI and PHOG in a one-versus-eight RF classifier. The improvement performance yields the significance of applying hierarchal system.

Moreover, to further evaluate the classification performance, confusion matrix is used to visualise the discrepancy between the actual class labels and predicted




(a) Bar plot



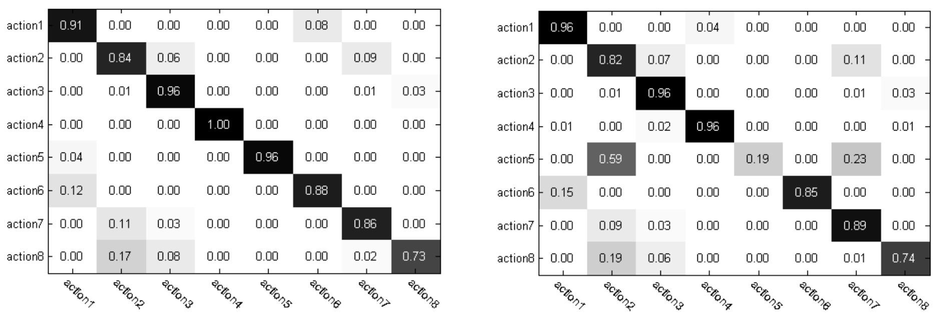

(b) Box plot

Fig. 18. Plot of experiment result in the hierarchal system

results from the classification. Confusion matrix gives the full picture at the errors made by a classification model. The confusion matrix shows how the predictions are made by the model. The rows correspond to the known class of the data, that is, the labels in the data. The columns correspond to the predictions made by the model.The value of each of element in the matrix is the number of predictions made with the class corresponding to the column. For example,with the correct value as represented by the row. Thus, the diagonal elements show the number of correct classifications made for each class, and the off-diagonal elements show the errors made. The confusion matrices of the hierarchal system and non-hierarchal system are shown in Fig. 19(a) and Fig. 19(b), respectively. In Fig. 19(b), the accuracy of action 5 is only 19% and the action 5 is confused into action 2 with a rate of 59%

Table 3. Classification Accuracy

| | **Classification Accuracy(%)** | | | |
|---|---|---|---|---|
| | **KNN** | **RF** | **SVM** | **MLP** |
| Level one | 99.27 | **99.87** | 99.87 | 99.87 |
| Level two-no-sg | 78.68 | **91.02** | 88.86 | 80.32 |
| Level two-sg | 79.63 | **99.48** | 94.18 | 88.20 |
| Level three | 100 | 99.83 | **100** | 100 |
| Level four | 95.60 | 93.60 | **96.00** | 83.40 |
| Hierarchal | **89.62** | | | |
| No hierarchal | 70.47 | **88.57** | 84.31 | 76.22 |



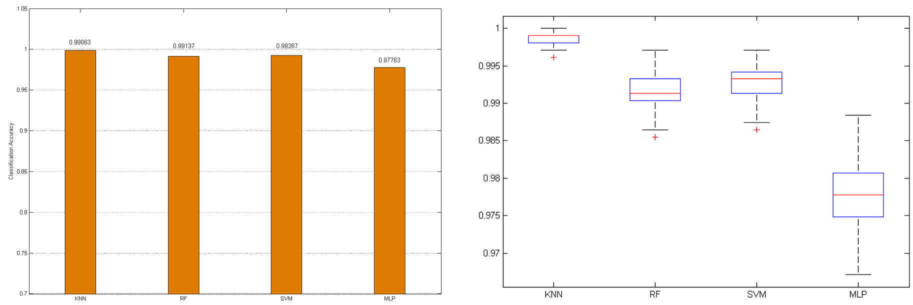

(a) Using hierarchal system (b) Without hierarchal system

Fig. 19. Confusion matrix

and action 7 with a rate of 23%. However,as shown in Fig. 19(a), the accuracy of action 5 is increased to 96% which means that 77% subsets of action 5 is closer to the others class in a non-hierarchal system by the feature of GEI.

### 8.2. *Dangerous Behaviour Classification Performance*

From the motion clips belonging to the classes of *IntHd* and *IntHd.DB*, we extracted about 10 thousand frames. We manually labeled these 10 thousand frames into two classes, one is *No Hand in Profile* and the other is *Hand in Profile*, as illustrated in Fig. 17. We setup a holdout experiment based on randomly dividing the 10 thousand frames into a training dataset (90% of the 10 thousand feature vectors extracted from the 10 thousand frames) and a test dataset (10% of the 10 thousand feature vectors extracted from the 10 thousand frames). Using the holdout experiment approach, only the test dataset is used to estimate the generalisation error. We repeat the holdout experiment 100 times by randomly splitting the 10 thousand features and recorded the classification results. The bar plot and box plot shows




(a) Bar plot (b) Box plot

Fig. 20. Experiment result in the dangerous behaviour classification

Table 4. Confusion matrix for the result from KNN classifier.
(I) No Hand in Profile,(II) Hand in Profile

| **class** | **I** | **II** |
|---|---|---|
| **I** | 99.9% | 0.1% |
| **II** | 0 | 100% |

the classification performance among four commonly used classifiers in Fig. 20(a) and Fig. 20(b). The result of classification rate of KNN, RF, SVM and MLP are 99.86%, 99.14%, 99.27% and 97.76%, respectively. The box plot in Fig. 20(b) further verifies that KNN classifier offers the best classification performance rate of the four classifiers. The confusion matrix of KNN shown in table 4 indicates that only 0.1% of class I samples are misclassified into class II while all class II samples are correctly classified.

### 8.3. *Comparison one - test/train data size ratio*

The SEU driving dataset contains 20 videos. In this subsection, three groups of hierarchal classification holdout experiment are conducted using different test/training data size ratios, each of which uses 20%, 30% and 40% of the dataset as testing data, respectively. Based on the best result reported in the section 8.1, RF classifier is used in previous two levels while SVM classifier is used in last two levels. In each group of the holdout experiments, specified proportion of the test videos are selected randomly each time, while the remaining videos are used for training. Each group of holdout experiments is repeated 100 times. Fig. 21 shows the overall average accuracies, which are compared with the default experiment that uses 10% of data for testing in section 8.1. The comparison result demonstrated that the variance in our model parameters estimation and the testing performance statistic

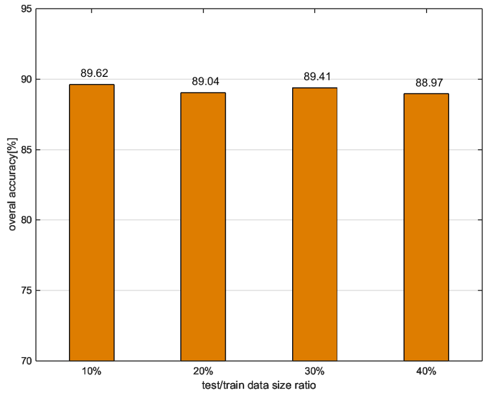



Fig. 21. The accuracy comparison of hierarchal classification experiment using different test/train data size ratios

.

is acceptable.

### 8.4. *Comparison two - best reported results with other approaches*

We treat the driver behaviour as spatio-temporal actions instead of static images [49,12,60,61]. Firstly, We hierarchically recognise motion clips under the framework of the action recognition. Then, based on the prior knowledge of categories of dangerous behaviour (eg. eating, smoking and responding a cellphone call), we further classify the head-related motion clip by the combination of PHOG and KNN, achieving a high accuracy of 99.86%. In our hierarchal classification system, we achieve 96.36% accuracy rate for class 5 (*IntHd*) and 88.41% accuracy rate for class 6 (*IntHd.DB*). We roughly estimate the responding cell phone recognition accuracy as $(96.36\% + 88.41\%)/2 * 100\% = 92.39\%$, and eating/smoking recognition accuracy as $(96.36\% + 88.41\%)/2 * 99.9\% = 92.29\%$.

Table 5. Classification Accuracy compared with other six approaches

| | Operating Shift Gear | Eating/Smoking | Responding a cellphone |
|---|---|---|---|
| Baseline[60] | 89.66 | 86.96 | 88.38 |
| MWT+MLP[61] | **92.82** | 87.59 | 83.01 |
| Proposed | 91.37 | **92.39** | **92.29** |

To provide a comprehensive performance evaluation and due to the different design schema (image classification v.s. time-serious image sequence classification), the best reported results are used to compare with two pervious approaches that using SEU dataset including (i) the method proposed in [60], which represents the posture pattern by contourlet transform on skin region, and (ii) the method proposed in [61], which extracts feature using mutiwavelet transform method from skin region. From the Table 5, our approach outperforms other approaches in dangerous driver behaviours recognition including responding a cellphone call ,eating and smoking.

### 8.5. *Discussion*

We apply gait energy image representation combined with shifting of ROI, skin region analysis and projection histogram in different levels of our hierarchal classification system which proves: (i)improved overall performance(89.62%) compared to traditional flat classification(88.57%) and (ii)classification accuracy for each class increases to no less than 73%. The hierarchy of the system and the representation feature used in each hierarchy can be further improved in later extension of our work. In addition, we combined PHOG and KNN in the classification of dangerous behaviours, which resulted in a high recognition rate of 99.86%. But eating and smoking are very similar behaviours and they are difficult to distinguish. They are labeled as the same class in our work. Further extension work is suggested to explore a better solution to distinguish eating and smoking.

## 9. Conclusion

This paper addresses the importance of automatic understanding and characterisation of driver behaviours in preventing motor vehicle accidents and presents a novel system for vision-based driver behaviour recognition. We verify our approach on the SEU driving dataset which includes activities of normal driving, responding to a cell phone call, eating and smoking. After pre-processing for illumination variations and motion clip segmentation, eight classes of behaviours are extracted for classification. By joint application of gait energy image, pyramid histogram of oriented gradients, hand skin-region segmentation and the hierarchal classification, our overall accuracy is over 89.62%. While there is an overall accuracy increase of 1.05% when compared to non-hierarchal classification system, the individual classification accuracy for each class increases to no less than 73%. We also estimate two dangerous driver behaviour, that is *IntHd.phone* and *IntHdDB.eat*, with an overall recognition rate of 91.87%.

## References

1. A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics* **9** (Jan 1979) 62–66.
2. Y. Bao-Cai, F. Xiao and S. Yan-Feng, Multiscale dynamic features based driver fatigue detection, *International Journal of Pattern Recognition and Artificial Intelligence* **23**(3) (2009) 575–589.

3. L. Bergasa, J. Nuevo, M. Sotelo, R. Barea and M. Lopez, Real-time system for monitoring driver vigilance, *IEEE Transactions on Intelligent Transportation Systems* **7** (March 2006) 63–77.
4. A. Bobick and J. Davis, The recognition of human movement using temporal templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23** (Mar 2001) 257–267.
5. A. Bosch, A. Zisserman and X. Munoz, Representing shape with a spatial pyramid kernel, in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, *CIVR '07* (ACM, New York, NY, USA, 2007) pp. 401–408.
6. G. Bradski and J. Davis, Motion segmentation and pose recognition with motion history gradients, in *Fifth IEEE Workshop on Applications of Computer Vision* (December 2000) pp. 238–244.
7. M. Brand and V. Kettnaker, Discovery and segmentation of activities in video, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** (Aug 2000) 844–851.
8. L. Breiman, Random forests, *Machine Learning* **45**(1) (2001) 5–32.
9. A. Cheddad, J. Condell, K. Curran and P. M. Kevitt, A skin tone detection algorithm for an adaptive approach to steganography, *Signal Processing* **89**(12) (2009) 2465–2478, Special Section: Visual Information Analysis for Security.
10. F.-C. Cheng, S.-C. Huang and S.-J. Ruan, Illumination-sensitive background modeling approach for accurate moving object detection, *IEEE Transactions on Broadcasting* **57** (Dec 2011) 794–801.
11. S. Y. Cheng, S. Park and M. M. Trivedi, Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis, *Computer Vision and Image Understanding* **106**(23) (2007) 245–257.
12. S. Cheng and M. Trivedi, Vision-based infotainment user determination by hand recognition for driver assistance, *IEEE Transactions on Intelligent Transportation Systems* **11** (Sept 2010) 759–764.
13. J. Choi, H. J. Chang, Y. J. Yoo and J. Y. Choi, Robust moving object detection against fast illumination change, *Computer Vision and Image Understanding* **116**(2) (2012) 179–193.
14. L. Chunli and W. KeJun, A behavior classification based on enhanced gait energy image, in *2010 2nd International Conference on Networking and Digital Society*, Vol. 2 (May 2010) pp. 589–592.
15. M. Davy, F. Desobry, A. Gretton and C. Doncarli, An online support vector machine for abnormal events detection, *Signal Processing* **86**(8) (2006) 2009–2025, Special Section: Advances in Signal Processing-assisted Cross-layer Designs.
16. D. Demirdjian and C. Varri, Driver pose estimation with 3d time-of-flight sensor, in *IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems* (March 2009) pp. 16–22.
17. Y. Dong, Z. Hu, K. Uchimura and N. Murayama, Driver inattention monitoring system for intelligent vehicles: A review, *Intelligent Transportation Systems, IEEE Transactions on* **12** (June 2011) 596–614.
18. A. Doshi and M. Trivedi, On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes, *IEEE Transactions on Intelligent Transportation Systems* **10** (Sept 2009) 453–462.
19. M.-P. Dubuisson and A. K. Jain, Contour extraction of moving objects in complex outdoor scenes, *International Journal of Computer Vision* **14**(1) (1995) 83–105.
20. X. Fan, B.-C. Yin and Y.-F. Sun, Yawning detection for monitoring driver fatigue, in *2007 International Conference on Machine Learning and Cybernetics*, Vol. 2 (Aug 2007) pp. 664–668.

21. Z. Feng and T.-J. Cham, Video-based human action classi.cation with ambiguous correspondences, in *Computer Vision and Pattern Recognition - Workshops* (June 2005) pp. 82–82.
22. L. Gorelick, M. Blank, E. Shechtman, M. Irani and R. Basri, Actions as space-time shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29** (Dec 2007) 2247–2253.
23. J.-E. Ha and W.-H. Lee, Foreground objects detection using multiple difference images, *Optical Engineering* **49**(4) (2010).
24. J. Han and B. Bhanu, Individual recognition using gait energy image, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** (Feb 2006) 316–322.
25. Q. Ji, Z. Zhu and P. Lan, Real-time nonintrusive monitoring and prediction of driver fatigue, *IEEE Transactions on Vehicular Technology* **53** (July 2004) 1052–1068.
26. T. Kato, T. Fujii and M. Tanimoto, Detection of driver's posture in the car by using far infrared camera, in *Proceedings of the IEEE Intelligent Vehicles Symposium* (June 2004) pp. 339–344.
27. Y. Ke, R. Sukthankar and M. Hebert, Event detection in crowded videos, in *IEEE 11th International Conference on Computer Vision* (Oct 2007) pp. 1–8.
28. V. Kecman, *Learning and soft computing [electronic book] : support vector machines, neural networks, and fuzzy logic models / Vojislav Kecman.*Complex adaptive systems, Complex adaptive systems (Cambridge, Mass. : MIT Press, 2001., 2001).
29. I. Laptev, B. Caputo, C. Schldt and T. Lindeberg, Local velocity-adapted motion events for spatio-temporal recognition, *Computer Vision and Image Understanding* **108**(3) (2007) 207 – 229, Special Issue on Spatiotemporal Coherence for Visual Motion Analysis.
30. D.-S. Lee, Effective gaussian mixture learning for video background subtraction, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (May 2005) 827–832.
31. H.-W. Lin, J.-L. Wu and M.-C. Hu, *Gait-based action recognition via accelerated minimum incremental coding length classifier*, Lecture Notes in Computer Science, Vol. 7131 LNCS, 2012).
32. C. Liu, F. Chang and Z. Chen, Rapid multiclass traffic sign detection in high-resolution images, *Intelligent Transportation Systems, IEEE Transactions on* **15** (Dec 2014) 2394–2403.
33. O. Masoud and N. Papanikolopoulos, A method for human action recognition, *Image and Vision Computing* **21**(8) (2003) 729–743.
34. E. Murphy-Chutorian and M. Trivedi, Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness, *IEEE Transactions on Intelligent Transportation Systems* **11** (June 2010) 300–311.
35. C. N., S. Jr. and A. A. Freitas, A survey of hierarchical classification across different application domains, *Data Mining and Knowledge Discovery* **22**(1-2) (2011) 31–72.
36. J. A. Nasiri, N. M. Charkari and K. Mozafari, Energy-based model of least squares twin support vector machines for human action recognition, *Signal Processing* **104**(0) (2014) 248 – 257.
37. Online, Who world report on road traffic injury prevention (2004), *http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/ .*
38. *Online, Traffic safety facts 2012: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system (2012), http://www-nrd.nhtsa.dot.gov/Pubs/812032.pdf .*
39. *Online, Transportation forecast: Light duty vehicles (2014),*

*http://www.navigantresearch.com/research/transportation-forecast-light-duty-vehicles .*

40. *P. Peursum, H. Bui, S. Venkatesh and G. West, Human action segmentation via controlled use of missing data in hmms, in* 17th International Conference on Pattern Recognition*, Vol. 4 (Aug 2004) pp. 440–445 Vol.4.*
41. *M. Piccardi, Background subtraction techniques: a review, in* Systems, Man and Cybernetics, 2004 IEEE International Conference on*, Vol. 4 (Oct 2004) pp. 3099–3104 vol.4.*
42. *C. Rao, A. Yilmaz and M. Shah, View-invariant representation and recognition of actions,* International Journal of Computer Vision ***50****(2) (2002) 203–226.*
43. *H. Sahbi and D. Geman, A hierarchy of support vector machines for pattern detection,* Journal of Machine Learning Research ***7*** *(2006) 2087–2123.*
44. *J. Schmudderich, V. Willert, J. Eggert, S. Rebhan, C. Goerick, G. Sagerer and E. Korner, Estimating object proper motion using optical flow, kinematics, and depth information,* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on ***38*** *(Aug 2008) 1139–1151.*
45. *W. R. Tan, C. S. Chan, P. Yogarajah and J. Condell, A fusion approach for efficient human skin detection,* IEEE Transactions on Industrial Informatics ***8*** *(Feb 2012) 138–147.*
46. *C. Tran, A. Doshi and M. M. Trivedi, Modeling and prediction of driver behavior by foot gesture analysis,* Computer Vision and Image Understanding ***116****(3) (2012) 435–445.*
47. *C. Tran and M. Trivedi, Towards a vision-based system exploring 3d driver posture dynamics for driver assistance: Issues and possibilities, in* Proceedings of the IEEE Intelligent Vehicles Symposium *(June 2010) pp. 179–184.*
48. *M. Trivedi, T. Gandhi and J. McCall, Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety,* Intelligent Transportation Systems, IEEE Transactions on ***8*** *(March 2007) 108–120.*
49. *H. Veeraraghavan, N. Bird, S. Atev and N. Papanikolopoulos, Classifiers for driver activity monitoring,* Transportation Research Part C: Emerging Technologies ***15****(1) (2007) 51–67.*
50. *S. Vitaladevuni, V. Kellokumpu and L. Davis, Action recognition using ballistic dynamics, in* IEEE Conference on Computer Vision and Pattern Recognition *(June 2008) pp. 1–8.*
51. *E. Wahlstrom, O. Masoud and N. Papanikolopoulos, Vision-based methods for driver monitoring, in* Proceedings of the IEEE Intelligent Transportation Systems*, Vol. 2 (Octuber 2003) pp. 903–908.*
52. *P. Watta, S. Lakshmanan and Y. Hou, Nonparametric approaches for estimating driver pose,* IEEE Transactions on Vehicular Technology ***56*** *(July 2007) 2028–2041.*
53. *D. Weinland, R. Ronfard and E. Boyer, A survey of vision-based methods for action representation, segmentation and recognition,* Computer Vision and Image Understanding ***115****(2) (2011) 224–241.*
54. *T. Whytock, A. Belyaev and N. Robertson,* Improving robustness and precision in GEI + HOG action recognition*Lecture Notes in Computer Science, Lecture Notes in Computer Science 2013.*
55. *D. Wu and L. Shao, Silhouette analysis-based action recognition via exploiting human poses,* IEEE Transactions on Circuits and Systems for Video Technology ***23*** *(Feb 2013) 236–243.*
56. *J. xiong Dong, L. Devroye and C. Suen, Fast svm training algorithm with decomposition on very large data sets,* IEEE Transactions on Pattern Analysis and Machine

Intelligence ***27*** *(April 2005) 603–618.*

57. *C. Yan, B. Zhang and F. Coenen, Driving posture recognition by joint application of motion history image and pyramid histogram of oriented gradients.,* International Journal of Vehicular Technology ***2014*** *(2014).*
58. *H. Yi, D. Rajan and L.-T. Chia, A new motion histogram to index motion content in video segments,* Pattern Recognition Letters ***26****(9) (2005) 1221–1231.*
59. *X. Zhang, N. Zheng, F. Mu and Y. He, Head pose estimation using isophote features for driver assistance systems, in* Proceedings of the IEEE Intelligent Vehicles Symposium *(June 2009) pp. 568–572.*
60. *C. Zhao, B. Zhang, J. He and J. Lian, Recognition of driving postures by contourlet transform and random forests,* Intelligent Transport Systems, IET ***6*** *(June 2012) 161–168.*
61. *C. Zhao, Y. Gao, J. He and J. Lian, Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier,* Engineering Applications of Artificial Intelligence ***25****(8) (2012) 1677 – 1686.*
62. *A. Zimek, F. Buchwald, E. Frank and S. Kramer, A study of hierarchical and flat classification of proteins,* Computational Biology and Bioinformatics, IEEE/ACM Transactions on ***7*** *(July 2010) 563–571.*
63. *X. Zou and B. Bhanu, Human activity classification based on gait energy image and coevolutionary genetic programming, in* 18th International Conference on Pattern Recognition*, Vol. 3 (Aug 2006) pp. 556–559.*